A comparison of manual anthropometric measurements with Kinect-based scanned measurements in terms of precision and reliability

Sara Bragança[1], Pedro Arezes[2], Miguel Carvalho[3], Susan P. Ashdown[4], Ignacio Castellucci[5], Celina Leão[2]

[1]Research and Innovation, Southampton Solent University, Southampton, United Kingdom

[2]Department of Production and Systems, University of Minho, 4800-019, Guimarães, Portugal

[3]Department of Textile Engineering, University of Minho, 4800-019, Guimarães, Portugal

[4]Department of Fiber Science & Apparel Design, Cornell University, Ithaca, NY, USA

[5]Escuela de Kinesiología, Facultad de Medicina, Universidad de Valparaíso, Valparaíso, Chile

Abstract

Background Collecting anthropometric data for real-life applications demands a high degree of precision and reliability. It is important to test new equipment that will be used for data collection.

Objective Compare two anthropometric data gathering techniques – manual methods and a Kinect-based 3D body scanner – to understand which of them gives more precise and reliable results.

Methods The data was collected using a measuring tape and a Kinect-based 3D body scanner. It was evaluated in terms of precision by considering the regular and relative Technical Error of Measurement and in terms of reliability by using the Intraclass Correlation Coefficient, Reliability Coefficient, Standard Error of Measurement and Coefficient of Variation.

Results The results obtained showed that both methods presented better results for reliability than for precision. Both methods showed relatively good results for these two variables, however, manual methods had better results for some body measurements.

Conclusion Despite being considered sufficiently precise and reliable for certain applications (e.g. apparel industry), the 3D scanner tested showed, for almost every anthropometric measurement, a different result than the manual technique. Many companies design their products based on data obtained from 3D scanners, hence, understanding the precision and reliability of the equipment used is essential to obtain feasible results.

Keywords: anthropometry; 3D body scanner; repeatability; body measurements

1. Introduction

When studies require the gathering of anthropometric data for the design of new products it is very important to ensure that the results appropriately reflect the characteristics of the studied population [1,2]. The importance of precision, reliability and validity of anthropometric data has been frequently studied [3,4]. However, reports on physical measurements in human populations frequently do not include estimates of measurement errors [5]. Reliability and the appropriate representation of the reality is also crucial is when using new measuring techniques and equipment. In this case it very important that the results obtained are close to the real value and are similar to the already proven methods. Most importantly, whatever the application or method used, the measurement of the human form needs to be practical and accurate [6,7].

Until recently, anthropometric measurements have been limited to traditional manual techniques, using anthropometers, calipers and measuring tapes. These techniques are simple to use and inexpensive but they also have some inherent limitations. Among these are: need for careful equipment calibration and trained observers; time-consuming nature of multiple measurement acquisition; and participants' compliance [8,9].

According to Geeta et al. [10], traditional anthropometry is usually the most error prone and with the lowest correlation coefficients when compared with other measuring techniques. Studies have shown that if the measurements are taken by different observers there can be differences of over 33mm and that differences of this order may not be detected with a measuring tape, especially in girth measurements [11,12].

Nowadays, these limitations are being overcome by the use of anthropometric measuring systems that use computer-based techniques, like three-dimensional body scanners. Body scanners proved to be an efficient and versatile equipment that is less time-consuming and invasive than hand-measurements [13,14]. The data obtained can be stored and used as a reference for changes in shape and size of the human body in life-long studies or for many other applications ranging from determining garment sizes for the clothing industry [15,16] to researching anthropometric measurements to identify patients with scoliosis or other diseases [17]. Nevertheless, these systems can also be error prone and must be evaluated and validated before a substantial use [18–20]. The most important evaluation criterion for all new measurement technologies is the ability to obtain reliable, precise and accurate data [21].

There are many alternative body scanning systems using a variety of technologies, wich include (i) 2D silhouette images converted to 3D models, (ii) white light phase based image capture, (iii) laser-based image capture, and (iv) radio-wave linear array image capture [22,23]. More recently systems have been developed that use infrared light sources. Table 1 briefly describes the various types of technology available for 3D body scanners.

Table 1. Types of technology used by 3D body scanners (Adapted from Daanen and Haar (27)).

Type of technology	**Description**
Laser line systems	A laser line, projected on the body from all sides, is viewed by cameras that are triangulated at a fixed angle. The advantage of a single line is that the sensor can easily detect it and very accurately compute how the projected 2D line is deformed over the 3D surface. The sequential captured 3D lines (generally taken at 1 or 2 mm increments) are then merged to form the complete 3D image.
Structured light systems / Photogrammetry	A structured light system projects a full structured light pattern into the surface of the body from the front and from the back and, from the sensed deformed pattern, a full 3D image is calculated. This light pattern can consist of dots, bars, or any other patterns. There is a specific type of structured light scanner that uses a pattern of near-infrared, invisible for the human eye, but visible for its sensor. The advantage of a structured light scanner is the speed of capturing data from the full body. Structured light scanning is so fast, that it can be used for 4D scanning; i.e. real time 3D scanning at 200 Hz. This offers opportunities to couple movement registration to 3D shape analysis.
Multi-view camera system	A 3D image is acquired from two or more cameras. A stereo-camera records two images at the same time from a different viewpoint. From the content in the two images the depth to the body can be calculated and converted into a dense 3D image in real-time. The advantage of a stereo-camera system is that no laser line or light pattern is transmitted, which means that environmental light cannot interfere with the pattern. However, using the line or patterns enables a 3D image with higher resolution and accuracy.
Millimeter waves	There are active and passive millimeter wave scanners. Active scanners use the reflection patterns of millimeter waves projected on the body. Passive scanners process the millimeter waves that are emitted by the human skin. Millimeter waves offer the advantage that they pass through most clothes but not the skin. Thus, the shape of the body can be captured without undressing. This offers an advantage in time and effort, but may introduce an ethical problem because the private parts of the participants can be seen. Millimeter wave scanners are currently employed at airports for the detection of metal parts under garments and offer an alternative for low radiation x-ray scanners.

Body scanning systems normally consist of one or more light sources, one or more vision or capturing devices, software, computer systems and monitor screens to visualize the data capture process [24]. The major scan technologies in use are those using laser and white light, but infrared systems (a specific type of structured light systems) have recently become popular (e.g. TC2, Size Stream, Kinect). It is possible to find many studies that compared the various types of existing body scanners. Jones and Rioux; Olds & Honey; Daanen and Haar; and Daanen and van de Water [24–27] discuss the use of 3D whole body scanners in anthropometry overall, giving a good overview of the evolution of body scanning technology and the different scanners in use at the various times.

Microsoft Kinect is a relatively new device that, apart for gaming, can also be used as a 3D body scanner. However, due to the simplification of the skeleton it cannot be used for extremely accurate studies [28–30]. Nonetheless, it is possible to use it when there is no need for such accuracy, like in clothes or shoes sizing, indirect fat measurement or clinical rehabilitation [31,32]. Recent research has shown that the Kinect system is capable of creating 3D human models with similar accuracy to the expensive and complex 3D body scanning systems [31]. When compared to many traditional systems, the Kinect does not require markers to be placed on participants before data collection, which reduces the data collection time, but also reduces the data accuracy, bringing many errors (e.g. pose estimation [28,29]).

Regardless of the system used, it is crucial that the data collected does not contain errors and is reliable and precise. This paper focuses only on the evaluation of precision and reliability, by comparing two techniques for anthropometric data collection. Hence, the objective of this paper is to compare two anthropometric data gathering techniques – manual methods and a Kinect-based 3D body scanner – to understand which of them gives more precise and reliable results. With this information, it would be possible to conclude if the 3D technique could replace the manual technique or if they are both valid to obtain good results.

2. Precision and reliability in the literature

There are many ways of assessing the data collected and finding possible errors. In the literature, many the terms describe anthropometric measurement error: unreliability, imprecision, undependability, inaccuracy, lack of validity, lack of reliability, lack of reproducibility and bias [33]. However, the definitions and interpretation of each one of these terms vary from author to author. With the literature review it was possible to conclude that there is a great discrepancy between what the variables represent. Hence, it is important to clarify the meaning of these variable so that anthropometric surveys can be properly evaluated and compared. Nevertheless, the terms most used and that have more impact in the interpretation of the data are: precision, reliability and accuracy. Hereon, the definition provided will be based on some pertinent literature and on the authors' understanding.

Precision is the variability between repeated measures by the same observer or by different observers. It is a characteristic of a particular observer using a particular technique to measure a particular variable [34]. Imprecision can be caused by imperfections in the measuring equipment; improper training of the observers; or difficulties in the measurement of certain anthropometric characteristics (such as skinfolds or large circumferences) [5,35]. The greater the variability between repeated measurements, the greater the imprecision [33,34].

The imprecision of anthropometric measurements is often evaluated using the Technical Error of Measurement (TEM). TEM is the most common way to express the measurement error in anthropometry. The International Society for the Advancement of Kinanthropometry (ISAK) has also been using it as an evaluation index to the accreditation of new anthropometrists [10,36]. The TEM is the standard deviation between repeated measurements taken by a single observer or by multiple observers [34]. It is expressed in the same units as the variable measured and can be used to calculate confidence intervals [35]. Lower values of TEM indicate a higher precision of the measurements. Values above a certain limit are not considered acceptable. According to the ISAK manual [37], the acceptable TEM values should be 0.1kg for weight, 3cm for height and 2cm for girths. However, as the TEM changes according to age and other specific population characteristics, it can be difficult to determine the acceptable levels [34,35]. As such, to ease the comparison between different studies, different populations or different anthropometric measurements, the TEM is converted into a relative Technical Error of Measurement (%TEM). The more strict acceptable %TEM levels are 7.5% for skinfolds and 1.5% for others measurements [5,38]. This variable does not allow comparisons between studies that use more than two observers and, as such, Sicotte et al. [35] advised the use of the TEM instead in those cases.

Reliability is concerned to the extent to which the within-subject variability is caused by factors other than the variance of measurement error or physiological variation [10,33]. Measures of precision and reliability differ in their units and in the constraints on their possible values. Norton and Olds [34] explained another two differences: (i) precision is a characteristic of a particular measurer using a particular measurement technique on a particular variable, while reliability has the same features plus the additional feature of being depended on the variability of subjects, (ii) measures of precision may be used in subsequent calculations of, for example, confidence intervals, or of the sample size needed to satisfy certain criteria. Reliability can be divided in relative and absolute reliability [39–41]. Relative reliability relates to the consistency of the position of individuals in a group, i.e., the extent to which individuals maintain their position in a sample over repeated measurements. It can be quantified using correlation coefficients, such as the Intraclass Correlation Coefficient (ICC) and the Reliability Coefficient (R). Absolute reliability is associated with the consistency of scores of individuals, i.e., the degree to which repeated measurements vary for individuals. It can be quantified using the Standard Error of Measurement (SEM) and the Coefficient of Variation (CV).

ICC provides an estimate of reliability and it is an indicator of repeatability, since it allows assessing the agreement between two sets of data [42]. The results are unit-less and can be evaluated using a limit scale, ranging from 0 (less reliable) to 1 (more reliable) [8]. The interpretation of the results varies according to the author considered. Portney & Watkins [43], for example, used only a scale with two categories (values above or below 0.75). Another example is the use of a four point scale as in Trippolini et al. [44], where it is stated that ICC$\geq$0.90 are excellent; 0.75$\leq$ICC$\leq$0.90 good; 0.50$\leq$ICC$\leq$0.75 moderate; and ICC$\leq$0.50 poor. The largest classification comes from Geeta et al. [10], where the ICC is divided in six categories.

R indicates the proportion of between-subject variance free from measurement error. The results are also unit-less and also vary from 0 (all between-subject variation was caused by measurement error) to 1 (no measurement error) [5,35]. Higher values of R imply greater reliability and values larger than 0.95 are considered sufficiently reliable [5,33,35].

SEM is an absolute index of reliability and represents the expected standard deviation for a particular measurement [42]. The more reliable the measurement, the less error is observed around the mean [39]. The SEM has the same units as the measurements under consideration and the interpretation of the results focuses on the assessment of reliability within individual subjects [40]. Low SEM values represent greater reliability. Denegar & Ball [45] state that direct calculation of the SEM involves the determination of the standard deviation of a large number of data points from an individual.

CV expresses the sample variability relative to the mean. It gives a general impression of the method's performance. However, it can only be used for scales that include a true zero value and is not valid for interval scales (e.g. temperature data [46]). To better assess the results, they can be transformed into percentages, making them unit-independent. The results obtained can signify that the method has a good performance (≤5%) or that the method's performance is not very good (≥10%) [10]. The CV percentage is good indicator to use when comparing methods. Nonetheless, Bland [47] pointed out that expressing the error as a percentage, will make the smallest observation differ prominently from the largest observation. As such, Chinn [48] suggested that instead of CV, using ICC would be a better alternative since it relates the size of the error variation to the size of the variation of interest. Nevertheless, Atkinson & Nevill [49] argued that the CV should not be used to evaluate reliability, since there are other methods that are more appropriate.

Finally, accuracy can be defined as the extent to which the "true-value" of a measurement can be reached [50]. In turn, inaccuracy represents the systematic bias that reduces the chances of attaining the true value and can be caused by instrument error or by errors in the measurement technique [33]. However, many times, the accuracy analysis is compromised by the difficulty of obtaining the real "true-value".

3. Materials and methods

3.1. Participants

Thirty-seven volunteers, seventeen females and twenty males, aged from 18 to 30 years old, have been involved in this study. Table 2 shows the characteristics of the sample population studied (Mean ± Standard Deviation).

Table 2. Characterization of the sample.

	Females (n=17)	Males (n=20)	Total (n=37)
Age (years)	23.94 ± 3.33	24.10 ± 2.79	24.03 ± 3.01
Height (m)	1.66 ± 0.07	1.77 ± 0.06	1.72 ± 0.09
Weight (kg)	57.71 ± 7.61	74.77 ± 9.99	66.93 ± 12.36

The participants were personally invited for this study or via e-mail. Prior to the study all participants were informed of the detailed procedures and requirements of the test and signed an informed consent.

3.2. Materials

Before the trials, the process was explained to the participants. All the measurements were taken using two techniques: 3D and manual. The 3D measurements were recorded using a 3D body scanner based on Microsoft Kinect (with four Kinect for Windows 1.5 SDK sensors to acquire 3D data clouds of front and back surfaces of a whole body). Through precise registration of the images captured by each device, the front and back views of the body can be merged and reconstructed to form a smooth and complete digital model that is scalable and rotatable on the computer screen (Figure 1).

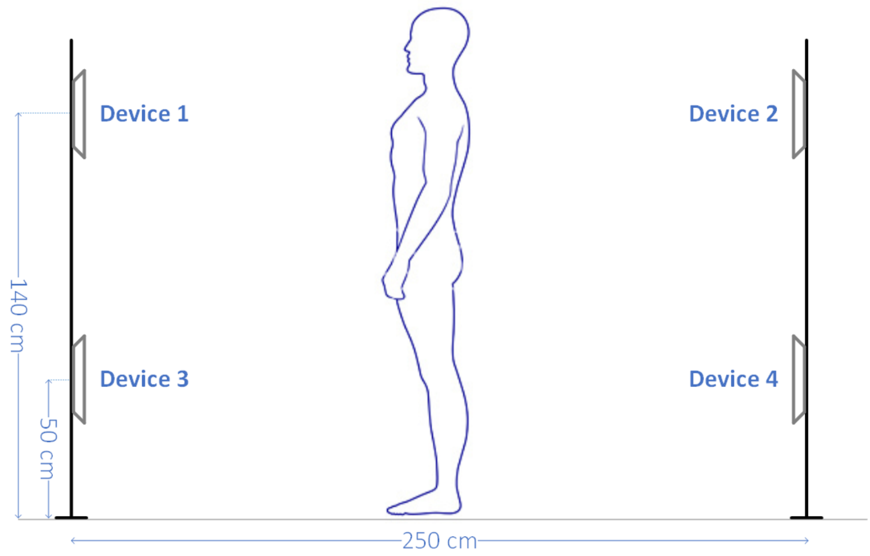

Figure 1. Configuration of the 3D body scanner for the measurement process.

KBI has a viewing volume of 2 m x 1 m x 0.8 m (height x width x depth), and outputs circumferences, lengths, heights and angles at many automatically-detected landmarks. It provides both hardware and software for 3D body imaging, which is quick (<4s imaging time), portable (on stands), compact (2mx1.5m footprint), relatively accurate and economical. The landmarks can be modified manually if they are not in accordance with visual assessments. The default automatic body dimensions that can be obtained with the automatic measurement extraction software (that comes with the body scanner) reflect the fact that this system was mainly developed for the clothing industry [51,52]. Furthermore, the measurement extraction software also allows the extraction of other measurements that are not predefined. This system was selected not only because it was the only body scanner available to the authors at the time but also due to the fact that this system had not yet been validated and tested in this way before.

A certified anthropometrist collected the manual measurements using a measuring tape. Before the manual data collection process, the body landmarks that would be needed for the study were identified, to minimize repetition error. The methods used for this were the ones described in the ISO 7250 [53] and the ISAK Manual [37].

During both measuring processes the participants wore their own underwear and assumed a stationary standing posture during the entire process (which was quicker with the 3D scanner).

3.3. Data collection

Two repetitions were performed for both techniques. The participants were first measured automatically (two scans; two files) and then manually with the ten body dimensions measured always in the same sequence. Between repetitions the participants were able to relax and go back to the measuring posture.

As analyzing all the possible body measurements would be an unrealistic task, a set of ten basic body measurements were selected for this study. Table 3 identifies the measurements collected and their definition according to the ISAK manual. The applicability of these definitions to the 3D scans was done as best as possible.

Table 3. Body measurements used in this study.

ID number	Measure	Definition according to the ISAK manual
1	Neck circumference	Measured immediately superior to the thyroid cartilage (Adam's apple)
2	Waist circumference	Measured at the level of the narrowest point between the lower costal (10th rib) border and the iliac crest
3	Hip circumference	Measured at the level of the greatest posterior protuberance of the buttocks
4	Mid-thigh circumference	Measured at the mid-trochanterion-tibiale-laterale site (linear distance between the trochanterion and tibiale laterale landmarks)
5	Knee circumference	Measured at the medial and lateral epicondyles of the femur
6	Calf circumference	Measured the most medial aspect of the calf at the level of the maximal girth
7	Ankle circumference	Measured at the narrowest point superior to the sphyrion tibiale
8	Shoulder width	Measured between the most lateral points on the acromion processes
9	Across chest length	Measured between the most lateral aspects of the thorax at the level of the mesosternale
10	Across back length	Measured between the most lateral aspects of the back at the level of the axillae

This selection includes some of the most significant different types of measurements: circumferences, widths, and lengths. All measurements recorded are surface body dimensions, not point to point linear measurements (such as those that would be taken with a caliper).

3.4. Data analysis

The obtained data was analyzed in terms of precision and reliability for both techniques. All the calculations took in consideration all the values (37 participants x 2 repetitions x 2 methods), not only the mean.

For the precision analysis the TEM and the %TEM were calculated (Equation 1 and Equation 2, respectively).

$$TEM = \sqrt{\frac{\sum D^2}{2N}} \quad \text{Equation 1}$$

$$\%TEM = \frac{TEM}{\bar{x}} \times 100 \quad \text{Equation 2}$$

D: the difference between the two repetitions; N: sample size; $\bar{x}$*: mean of the measurements*

The reliability analysis was twofold, calculating: (i) the relative reliability, with the ICC (two-way mixed model and absolute agreement type) and the R (Equation 3); and (ii) the absolute reliability, with the SEM and the CV (Equation 4 and Equation 5, respectively).

$$R = 1 - \left(\frac{TEM^2}{SD^2}\right) \quad \text{Equation 3}$$

$$SEM = SD \times \sqrt{1 - ICC} \quad \text{Equation 4}$$

$$CV = \frac{SD}{\bar{x}} \quad \text{Equation 5}$$

SD: standard deviation; $\bar{x}$*: mean of the measurements*

Additionally, Bland-Altman plots were constructed for the mean difference of each measurement for the two methods in order to determine the level of agreement between the two data type measurement techniques.

4. Results

The means and standard deviations of the measurements of all the 37 participants are presented in Table 4.

Table 4. Means and standard deviations of the measurements (in cm).

	Mean (Manual)	Mean (3D)	Standard Deviation (Manual)	Standard Deviation (3D)
1. Neck circumference	38,94	41,84	3,50	5,52
2. Waist circumference	75,40	81,75	9,43	10,42
3. Hip circumference	93,45	104,31	6,74	11,01
4. Mid thigh circumference	44,40	46,78	4,27	5,37
5. Knee circumference	36,56	37,21	2,59	4,07
6. Calf circumference	36,34	39,70	2,98	4,97
7. Ankle circumference	23,08	26,61	1,61	3,66
8. Shoulder width	45,41	46,98	4,51	5,75
9. Across chest length	36,20	36,00	3,09	5,44
10. Across back length	46,75	49,27	4,50	6,93

4.1. Precision

Table 5 presents the results of the TEM and %TEM calculation. TEM values were considered acceptable if lower than 2cm. For the %TEM, values higher than 1.5% were not considered acceptable and are marked with an asterisk. These thresholds were based on ISAK limits of acceptance for a level 1 certified anthropometrist [37].

Table 5. Results of the TEM (in cm) and %TEM calculations for the analyzed methods.

	TEM (Manual)	TEM (3D)	%TEM (Manual)	%TEM (3D)
1. Neck circumference	0.51	0.84	1.30	2.01*
2. Waist circumference	1.01	0.84	1.35	1.02
3. Hip circumference	0.99	1.06	1.06	1.02
4. Mid thigh circumference	0.48	1.01	1.08	2.15*
5. Knee circumference	0.26	0.33	0.70	0.90
6. Calf circumference	0.23	0.86	0.64	2.16*
7. Ankle circumference	0.25	0.56	1.08	2.09*
8. Shoulder width	0.66	1.25	1.44	2.66*
9. Across chest length	0.56	0.94	1.56*	2.62*
10. Across back length	0.64	1.21	1.37	2.46*

*not acceptable value

In terms of precision and according to the obtained results, all TEM values were lower than 2cm, meaning that there was some considerable precision. For the manual measurements, the lowest value (best result) occurred for the calf circumference (TEM_{Manual}=0.23), whilst the highest value (worst result) was recorded for the waist circumference (TEM_{Manual}=1.01). In the 3D measurements, the best score was for the knee circumference (TEM_{3D}=0.33) and the worst score was for the shoulder width (TEM_{3D}=1.25). Despite the fact that TEM calculations show that both the manual and the 3D method were precise, the values obtained with the manual technique were slightly lower than for the 3D method, indicating that there was more precision in this case.

Figure 2 depicts the differences between the results of the two analyzed methods. There are some measurements where the difference between methods was not very accentuated, such as the knee or hip circumferences. Opposing, there are some measurements that present differences higher than 0.5cm – mid-thigh circumference, calf circumference, shoulder width and across back length. The biggest difference occurs for the calf circumference, with the 3D technique recording the largest error (TEM_{Manual}=0.23; TEM_{3D}=0.86).

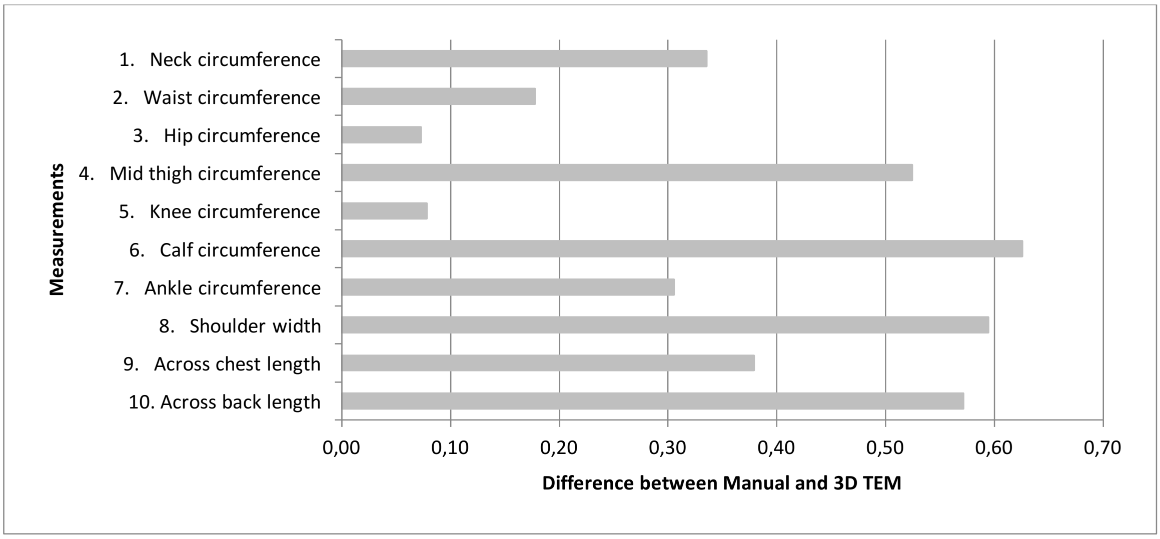

Figure 2. Difference between the TEM values of the manual method and the 3D method.

Regarding the %TEM, in manual measurements the best value recorded was for the calf circumference with 0.64% and the worst was 1.56% for the across chest length. In the scanner measurements the best value was 0.90% for the knee circumference and the highest %TEM value recorded was 2.66% for the shoulder width.

The results of the manual method show that only the across chest length recorded a value higher than 1.5%, whilst in the 3D method, there are seven measurements with values higher than this limit.

4.2. Reliability

The results of the first part of the relative reliability analysis, using the ICC, are presented in Table 6. The measurements with excellent and good reliability values for the ICC are marked with one or two asterisks, respectively (according to Trippolini et al. [44]).

Table 6. ICC results for the analyzed methods.

	ICC (Manual)	ICC (3D)
1. Neck circumference	0.80*	0.95**
2. Waist circumference	0.99**	0.99**
3. Hip circumference	0.98**	0.97**
4. Mid thigh circumference	0.99**	0.94**
5. Knee circumference	0.99**	0.98**
6. Calf circumference	0.99**	0.95**
7. Ankle circumference	0.98**	0.96**
8. Shoulder width	0.98**	0.91**
9. Across chest length	0.97**	0.94**
10. Across back length	0.98**	0.94**

*good; **excellent

Most of the ICC values indicate an excellent reliability. In the manual measurements, the results indicated excellent reliability in four cases (ICC_{Manual}=0.99). The lowest ICC value recorded in the manual measurements was 0.8 for the neck circumference, which still represents a good reliability. Surprisingly, the neck circumference was the only measurement where the ICC value was higher in the 3D technique. When comparing the two methods, all the measurements, except the neck circumference, present a slightly higher score in the manual technique, implying that it has a better reliability. In the 3D technique, the waist circumference was the measurement that ranked first, with the better result (ICC_{3D}=0.99) and the shoulder width was the measurement that ranked last with the worst result (ICC_{3D}=0.91).

The assessment of the relative reliability with the R (thresholds based on Arroyo et al. and Sicotte et al. [5,35]) showed that both methods present very similar results that were higher than 0.95. Almost all measurements have values of 0.98 or 0.99. In the cases where there was a small difference between methods, the manual method was always preferred, except for the hip circumference, where the 3D technique shows a better result (R_{Manual}=0.98; R_{3D}=0.99). The best scores (0.99), in the manual method were obtained for the waist, mid-thigh, knee and calf circumferences. In the 3D method, the best score of 0.99 was obtained for the waist, hip and knee circumferences. The worst scores were found for the across chest length (in the manual method with 0.97) and for the shoulder width (in the 3D method with 0.95). Table 7 shows the results of second part of the relative reliability analysis, using the R. For this variable, values higher than 0.95 were considered sufficiently reliable.

Table 7. Results of the R for the analyzed methods.

	R (Manual)	R (3D)
1. Neck circumference	0.98	0.98
2. Waist circumference	0.99	0.99
3. Hip circumference	0.98	0.99
4. Mid thigh circumference	0.99	0.96
5. Knee circumference	0.99	0.99
6. Calf circumference	0.99	0.97
7. Ankle circumference	0.98	0.98
8. Shoulder width	0.98	0.95
9. Across chest length	0.97	0.97
10. Across back length	0.98	0.97

Concerning absolute reliability, the SEM does not provide an assessment baseline, so, the lower the error, the higher the reliability. For the manual measurements, the highest value registered was 1.58cm for the neck circumference and the lowest value was 0.23cm for the calf circumference. In the 3D measurements case, the highest value registered 1.97cm for the hip circumference and the lowest value was 0.45 for the knee circumference. Again, the 3D method indicates less reliable results (higher values) for all measurements with the exception of the neck circumference, where it indicates a greater reliability (SEM_{Manual}=1.58; SEM_{3D}=1.18). In the manual method, the measurements knee circumference, calf circumference and ankle circumference were the ones with the lowest amount of error, all less than 0.5cm; whereas in the 3D method, most measurements have an error higher than 1cm. Table 8 shows the results of the SEM calculations. As SEM does not have a target value, the interpretation in this case was the lower the result, the higher the reliability.

Table 8. Results of the SEM (in cm) for the analyzed methods.

	SEM (Manual)	SEM (3D)
1. Neck circumference	1.58	1.18
2. Waist circumference	0.99	0.99
3. Hip circumference	0.98	1.97
4. Mid thigh circumference	0.47	1.28
5. Knee circumference	0.26	0.45
6. Calf circumference	0.23	1.17
7. Ankle circumference	0.25	0.70
8. Shoulder width	0.65	1.76
9. Across chest length	0.56	1.34
10. Across back length	0.64	1.65

CV results show that neither of the methods have a good performance since in all measurements the results were higher than the target value of 5%. In fact, all values in the 3D method and one value in the manual method were above the 10% margin, which indicates a not very good performance. In the manual technique, the best score was for the ankle circumference with 6.98%; while in the 3D techniques, the best score was 10.56% for the hip circumference. The worst score was recorded for the waist circumference in the manual method (CV_{Manual}=12.50%) and for the across chest length in the 3D method (CV_{3D}=15.12%). CV results are presented in Table 9. Values lower than 5% represent a good performance of the method (according to Geeta et al. [10]).

Table 9. Results of the CV for the analyzed methods.

	CV (Manual)	CV (3D)
1. Neck circumference	8.98%	13.19%
2. Waist circumference	12.50%	12.74%
3. Hip circumference	7.21%	10.56%
4. Mid thigh circumference	9.63%	11.48%
5. Knee circumference	7.08%	10.95%
6. Calf circumference	8.19%	12.52%
7. Ankle circumference	6.98%	13.76%
8. Shoulder width	9.94%	12.23%
9. Across chest length	8.53%	15.12%
10. Across back length	9.62%	14.06%

5. Discussion

After all the calculations, it was possible to verify that there were some measurements that were more precise and more reliable than others. Table 10 depicts the measurements that obtained the best and the worst scores for each variable analyzed.

Table 10. Best and worst results for each measurement for each evaluation parameter.

	Manual						3D					
	TEM	%TEM	ICC	R	SEM	CV	TEM	%TEM	ICC	R	SEM	CV
1. Neck circumference					**W**							
2. Waist circumference	**W**		**B**	**B**		**W**			**B**	**B**		
3. Hip circumference										**B**	**W**	**B**
4. Mid thigh circumference			**B**	**B**								
5. Knee circumference			**B**	**B**			**B**	**B**		**B**	**B**	
6. Calf circumference	**B**	**B**	**B**	**B**	**B**							
7. Ankle circumference						**B**						
8. Shoulder width							**W**	**W**	**W**	**W**		
9. Across chest length		**W**		**W**								**W**
10. Across back length												

B Best result; **W** Worst result

None of the measurements was the best or the worst in the two methods at the same time. While there are measurements that represent the best or the worst result in a given method, there was one measurement, the across back length, that was neither very reliable nor very precise in either method. The calf circumference is the only measurement that represents the best result in almost every evaluation parameter (excluding the CV) in the manual method. This situation can be understood because this was the easiest measurement to take manually, with a very well defined landmark.

The results show that, in general, the measurements presented better results for reliability than for precision. The scores obtained in the calculations of the ICC and R variables (relative reliability) allowed more than one measurement to be considered as the best result.

The ranking of each measurement inside each variable was plotted for both methods. The results of the calculation for each variable were placed in ascending or descending order according to the meaning of the value (e.g. in the TEM calculations, a higher value indicates a worst result whilst in the ICC calculation, a higher value indicates a better result). The ranking goes from 1 (better result) to 10 (worst result). Figure 3 shows this ranking for the manual method, with the values closer to the central axis more positive and those outside more negative.

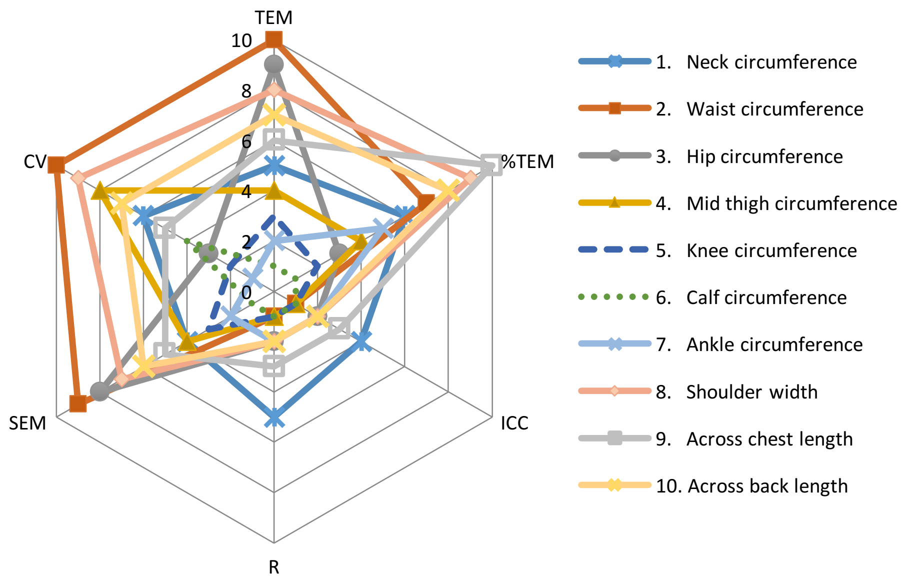

Figure 3. Measurement ranking within each variable for the manual technique.

This graphical representation helps to understand which measurements have the best and worst performances for each variable. As previously stated, the calf circumference was the measurement that demonstrates the best result for all variables, except CV. Opposite, it was possible to find the variable waist circumference and shoulder width, with most results ranked as the worst for almost all variables.

Accordingly, Figure 4 shows this ranking for the 3D method.

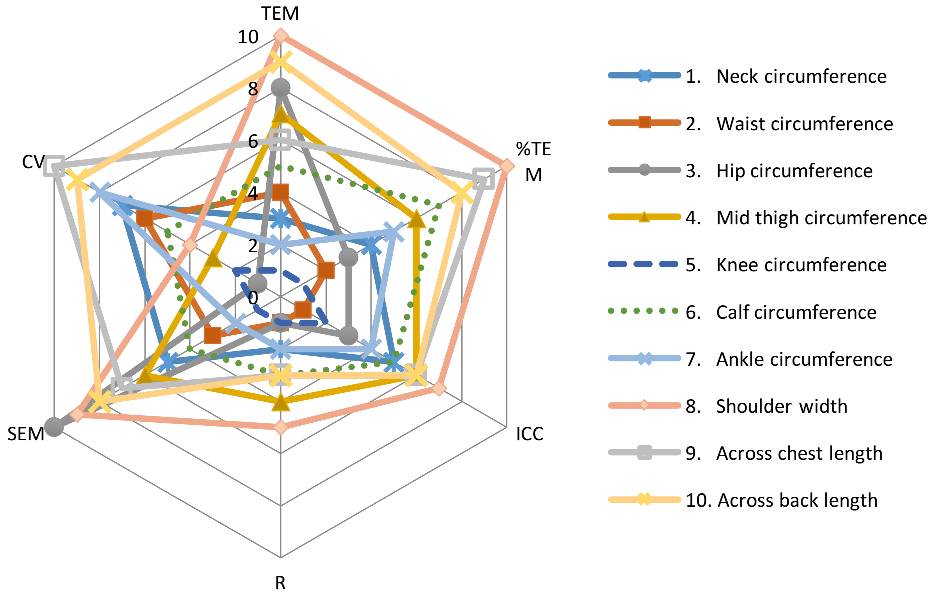

Figure 4. Measurement ranking within each variable for the 3D technique

Interestingly, in the 3D technique, the waist circumference measurement, that had one of the worst behaviors in the manual method, has now a better performance than many other measurements. Nonetheless, the shoulder width measurement still presents some of the worst results for almost every variable alongside with the across chest and across back lengths. This can be easily understood since in the 3D image of the scanner it was simple to identify the narrowest part of the body and its contour, whereas in the manual method the correct reading of the measurement can be influenced by the amount of body fat surrounding this area and by pressure exerted on the measuring tape. Areas with large amounts of accumulated fat tissue are more difficult to measure manually not only because of the difficulty to determine the correct pressure to exert but also because of the difficulty of identifying the correct landmark. Using manual techniques, palpation and recognition of some skeletal points are required to identify some landmarks. When those predefined locations are covered with excessive fat tissue the palpation process is very much complicated.

The several Bland-Altman plots were designed to illustrate the variability of the measurements, by showing the spread of differences in the two measurements of each body part, their mean difference and the upper and lower limits of agreement for both techniques. In all the measurements the mean difference value was very close to 0, which is a good indicator. However, this small difference indicates some evidence of systematic proportional bias (trend). The measurement that has a mean difference farthest from 0 was the hip circumference in the manual technique (Mean=0.4914cm). Figure 5 depicts the Bland-Altman plots for the hip circumference measurement for both techniques.

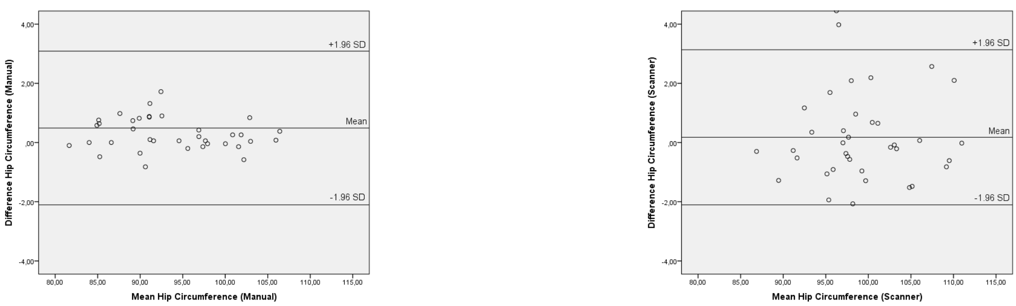

Figure 5. Bland-Altman plots for the hip circumference measurement (manual and 3D)

In terms of the limits of agreement, they are widely separated from the mean, which indicates some ambiguity in the results. In contrast, there are also some measurements where the limits of agreement were very close to the mean (knee circumference both manual and 3D; calf circumference manual and ankle circumference manual), which indicates that there was a smaller bias in these cases. An example of this small range in the limits of agreement can be found in Figure 6, for the knee circumference both in the manual and 3D techniques. These results are consistent with the previous calculations where these measurements where the ones that presented the best performance for almost every variable (see ranking of Figure 3 and Figure 4).

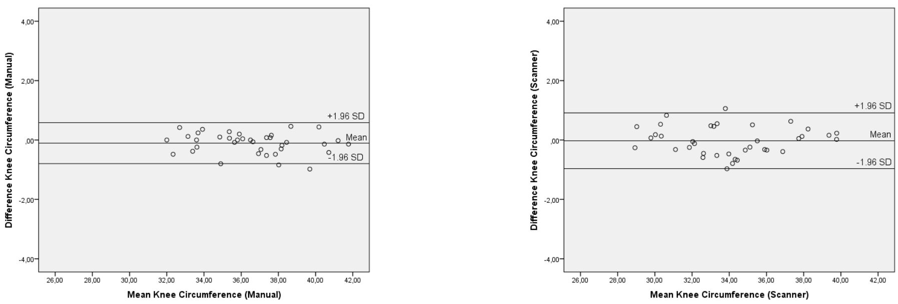

Figure 6. Bland-Altman plots for the knee circumference measurement (manual and 3D)

Regarding the dispersion of the dots, there are some measurements where there is some funneling effect, as the scatter around the bias line gets larger as the average gets higher. These measurements were all in the 3D method (neck, hip, the mid-thigh circumferences; shoulder width, across chest and across back length). In Figure 6 this can be noticed for the hip circumference (3D), where there is a greater variability of the data than in the manual method. Figure 7 also shows that the 3D method has more variability than the manual method.

Regardless of the values themselves, which represent a natural variation in measurements among the participants, it is also noticeable a difference on the variability of the measurements for body parts with different characteristics. The manual method is the one that is expected to have the largest variability due to natural configuration of fat and muscle. In the hip, the measurements are made over large areas of muscle mass and fat deposits (that would result in greater variation) whilst in the knee the measurements are made over bone, tendons and ligaments (allowing for closer values for the largest to the smallest person).

However, for both these examples the largest variability is registered in the 3D method. This can be explained by the different values obtained in each repetition. In fact, according to the results, the 3D method was the one that presented the worst results most of the time.

Although sometimes the identification of landmarks is easier to do in the 3D measurement software (as waist circumference), this identification can often be extremely challenging. Furthermore, the location of the landmarks on the participants' body may not be exactly the same in both methods, since the 3D measurement software automatically identifies the correct body locations. Despite being based on complex algorithms, this automatic identification sometimes presents certain errors. Hence, these imprecisions in the 3D technique can be due to the difficulty of identifying the proper landmark to perform the measurement on the computer screen. Figure 7 displays five repetitions of scans taken from one of the study participants. As it can be seen, the location of the several landmarks was not always the same.

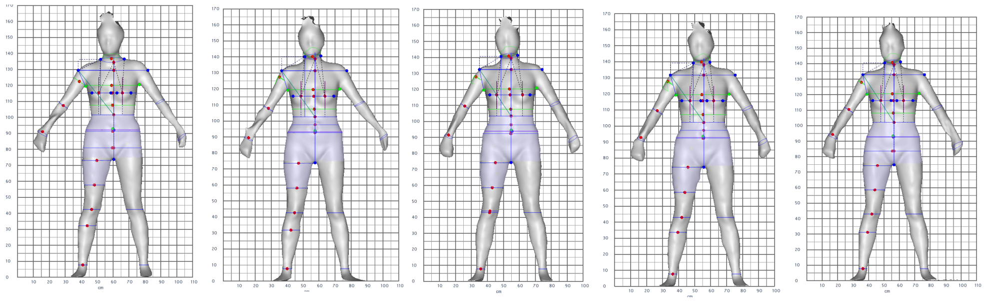

Figure 7. Problems in the automatic landmark identification by the 3D measurement software

For example in the calf region, it is possible to see that the blue line with the red dot is not placed exactly on the same spot in all five images. In the third repetition the calf was wrongly placed right below the knee (Figure 8 in yellow). The same problem can also be seen for the hip region. The landmark is only correctly placed in the first and last repetition, where on the others it is near the waist (Figure 8 in orange).

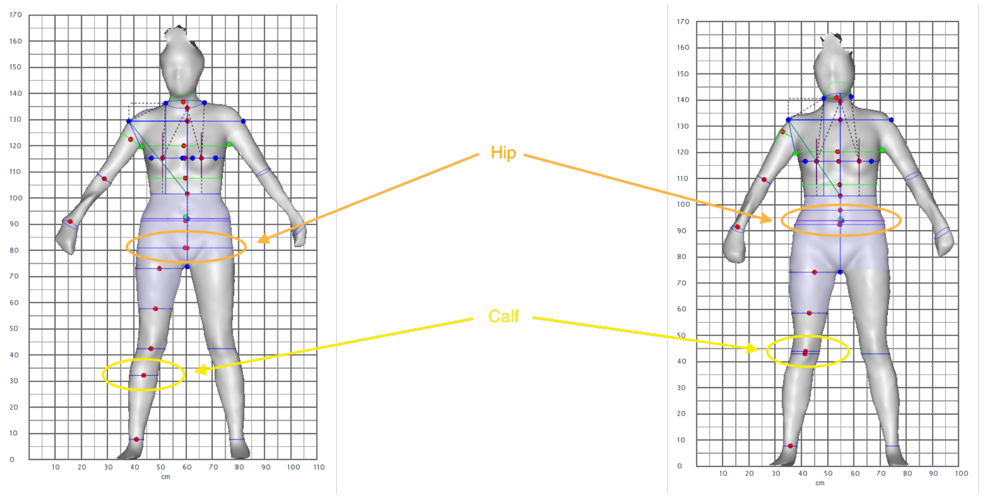

Figure 8. Detail of the problems in the automatic landmark identification in the knee and hip

Although this is an automatic process, it is possible to move the body location where the landmark was placed to another more suitable location. However, moving the landmark location empirically to the exact same location in the various repetitions is sometimes difficult and time-consuming. Nonetheless, modifying the landmark placement visually may provide the best compromise between time-consuming manual methods, and inconsistent automated scan measurements.

This fact, allied to other potential problems that may occur during a measuring session with a body scanner, if for example the participants move during the scanning process, causes the data to be less precise and reliable than the data obtained with manual methods.

These results seem to imply that the 3D technique is not very precise, since most of the measurements were above the limit value. However, this is the limit for highly trained human observers, set by ISAK (for high-level accredited personnel). Since this is a technique that involves little human interaction, perhaps the value of the method outperforms the manual method, and depending on the end use of the measurement, the limits should not be the same. If a less strict limit was imposed, as the novice ISAK anthropometrists (max %TEM=2%), the results obtained would still not be optimal, but they could be accepted. Nonetheless, for applications such as the clothing industry, where the requirement for level of precision of the measurements is not high, the data obtained through 3D methods is perfectly acceptable. In a previous study conducted by the authors, with the same data, the results obtained with the 3D scanner were compared to the ISO 20685 [54] and to other studies that used 3D scanners to collect anthropometric data. The study proved that this Kinect-based 3D scanner is comparable to other devices in the market and is able to give sufficiently good results. With a 3D capturing system it is possible to record many more measurements (virtually all possible measurements in the human body) than the necessary for a particular study. It is also useful for other studies where different measurements are taken in consideration, since the data is already stored in a computer and always accessible, without the need to measure the same participants in other body locations.

6. Limitations of the study

This study had some inherent limitations. The 3D scanning system based on the Kinect used for this study is not a commercially available, fully developed 3D body scanner, which may have affected the results obtained. Other body scanners specifically built for the purpose of anthropometric data collection might have shown better reliability and precision results.

Another limitation is with the automatic positioning of the landmarks in the 3D scans, which might have not been exactly the same used in the manual measurements. Moreover, as this study included two repetitions of the same measurement, moving the landmarks could also have affected the results.

The limits for all the error measurement evaluation calculations was originally meant for manual measurements. Perhaps, using these same limits for 3D scans is not the most appropriate assessment procedure. However, these are the only currently available option.

7. Conclusions

The results of this study showed that, for most of the selected body measurements, the manual method produced better results than the 3D body scanner. However, the difference between the two methods was not very large, as was observed with all the calculations. Hence, it can be said that for applications that require lower levels of precision and reliability (such as clothing, where one centimeter more or less does not have major impacts), the anthropometric data collected with the body scanner can be used.

As such, in future studies, where there is the possibility of using one of these methods, the selection of either one of them would be a valid option. Still, it should be pointed out that this study only encompassed the comparison of the manual methods with one particular body scanner, the only one that was available for the researchers, and that there may be other body scanners that could give worst or better results.

Scanners using white light and laser light sources have been on the market for a long time, unlike the relatively new infrared light source scanners, and may give better results. Also, the reliability and precision of the scanner output is directly related to the quality of the automated software that is used to extract measurements. Some studies on body scanners have been conducted over the years that conclude that they can be both reliable and precise enough for anthropometric data collection for many end uses. Using a 3D body scanner and in particular, the system based on the Microsoft Kinect, can be better than taking the measurements manually, as it is a much faster process with very similar results.

References

1. Hsiao H, Whitestone J, Kau T-Y, Hildreth B. Firefighter Hand Anthropometry and Structural Glove Sizing A New Perspective. Hum Factors J Hum Factors Ergon Soc. 2015;57(8):1359–77.
2. Mehrparvar AH, Mirmohammadi SJ, Hafezi R, Mostaghaci M, Davari MH. Static anthropometric dimensions in a population of Iranian high school students considering ethnic differences. Hum Factors J Hum Factors Ergon Soc. 2014;18720814549579.
3. Kouchi M, Mochimaru M, Bradtmiller B, Daanen H, Li P, Nacher B, et al. A protocol for evaluating the accuracy of 3D body scanners. Work. 2012;41(SUPPL.1):4010–7.

4. Okimoto MLLR, Ribeiro L, Klein AA. Photogrammetry procedures applied to anthropometry. Work. 2012;41(SUPPL.1):4046–52.

5. Arroyo M, Freire M, Ansotegui L, Rocandio A. Intraobserver error associated with anthropometric measurements made by dietitians. Nutr Hosp. 2010;25(6):1053–6.

6. Bretschneider T, Koop U, Schreiner V, Wenck H, Jaspers S. Validation of the body scanner as a measuring tool for a rapid quantification of body shape. Ski Res Technol. 2009;15(3):364–9.

7. Yoon JC, Radwin RG. The accuracy of consumer-made body measurements for women's mail-order clothing. Hum Factors J Hum Factors Ergon Soc. 1994;36(3):557–68.

8. Haniff J, Appannah G, Nor M, Safiza N, Wong N, Kee C, et al. Reliability and technical error of calf circumference and mid-half arm span measurements for nutritional status assessment of elderly persons in Malaysia. Malays J Nutr. 2008;14(2):137–50.

9. Wong J, Oh A, Ohta E, Hunt A, Rogers G, Mulliken J, et al. Validity and reliability of craniofacial anthropometric measurement of 3D digital photogrammetric images. Cleft Palate-Craniofacial J. 2008;45(3):232–9.

10. Geeta A, Jamaiyah H, Safiza M, Khor G, Kee C, Ahmad A, et al. Reliability, technical error of measurements and validity of instruments for nutritional status assessment of adults in Malaysia. Singapore Med J. 2009;50(10):1013–8.

11. Maylia E, Fairclough J, Nokes L, Jones M. Can thigh girth be measured accurately? A preliminary investigation. J Sport Rehabil. 1999;8(1):43–9.

12. Fairclough J, Mintowt-Czyz W, Mackie I, Nokes L. Abdominal girth: an unreliable measure of intra-abdominal bleeding. Injury. 1984;16(2):85–7.

13. Bougourd J, Dekker L, Grant Ross P, Ward J. A comparison of women's sizing by 3D electronic scanning and traditional anthropometry. J Text Inst. 2000;91(2):163–73.

14. Yu C-Y, Lo Y-H, Chiou W-K. The 3D scanner for measuring body surface area: a simplified calculation in the Chinese adult. Appl Ergon. 2003;34(3):273–8.

15. Szirovicza L, Ujević D, Drenovac M. The structure of body measurements for the determination of garment system for young Croatian men. Coll Antropol. 2002;26(1):187–97.

16. Istook C, Hwang S-J. 3D body scanning systems with application to the apparel industry. J Fash Mark Manag An Int J. 2001;5(2):120–32.

17. Schmitz A, Gäbel H, Weiss H, Schmitt O. [Anthropometric 3D-body scanning in idiopathic scoliosis]. Z Orthop Ihre Grenzgeb. 2001;140(6):632–6.

18. Kouchi M, Mochimaru M. Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry. Appl Ergon. 2011;42(3):518–27.

19. Robinette K, Daanen H. Precision of the CAESAR scan-extracted measurements. Appl Ergon. 2006;37(3):259–65.

20. Xu X, McGorry RW. The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures. Appl Ergon. 2015;49:47–54.

21. Fourie Z, Damstra J, Gerrits P, Ren Y. Accuracy and repeatability of anthropometric facial measurements using cone beam computed tomography. Cleft Palate-Craniofacial J. 2011;48(5):623–30.

22. Treleaven P, Wells J. 3D body scanning and healthcare applications. Computer (Long Beach Calif). 2007;40(7):28–34.

23. Istook CL, Hwang S. 3D body scanning systems with application to the apparel industry. J Fash Mark Manag. 2001;5(2):120–32.

24. Daanen HAM, van de Water GJ. Whole body scanners. Displays. 1998;19(3):111–20.

25. Jones PRM, Rioux M. Three-dimensional surface anthropometry: Applications to the human body. Opt Lasers Eng. 1997;28(1):89–117.

26. Olds T, Honey F. The use of 3D whole-body scanners in anthropometry. In: 9th International Conference of the International Society for the Advancement of Kinanthropometry. Routledge; 2006. p. 1–12.

27. Daanen HAM, Haar FB. 3D whole body scanners revisited. Displays. 2013;34(4):270–5.

28. Clark RA, Pua YH, Fortin K, Ritchie C, Webster KE, Denehy L, et al. Validity of the Microsoft Kinect for assessment of postural control. Gait Posture. 2012;36(3):372–7.

29. Ye M, Wang X, Yang R, Ren L, Pollefeys M. Accurate 3D pose estimation from a single depth image. In: IEEE International Conference on Computer Vision. 2011. p. 731–8.

30. Bragança S, Carvalho M, Xu B, Arezes P, Ashdown SP. A validation study of a Kinect based Body Imaging (KBI) device system based on ISO 20685:2010. In: 5th International Conference and Exhibition on 3D Body Scanning Technologies (3DBST). Proceedings of the Fifth International Conference and Exhibition on 3D Body Scanning Technologies (3DBST); 2014. p. 372–7.

31. Weiss A, Hirshberg D, Black MJ. Home 3D body scans from noisy image and range data. In: IEEE International Conference on Computer Vision. IEEE; 2011. p. 1951–8.

32. Fernández-Baena A, Susin A, Lligadas X. Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. In: IEEE International Conference on Intelligent Networking and Collaborative Systems. IEEE; 2012. p. 656–61.

33. Ulijaszek S, Kerr D. Anthropometric measurement error and the assessment of nutritional status. Br J Nutr. 1999;82(3):165–77.

34. Norton K, Olds T. Anthropometrica: a textbook of body measurement for sports and health courses. UNSW press; 1996.

35. Sicotte M, Ledoux M, Zunzunegui M-V, Aboubacrine S, Nguyen V-K. Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa. BMC Med Res Methodol. 2010;10(102):1–9.

36. Perini T, Oliveira G, Ornellas J, Oliveira F. Technical error of measurement in anthropometry. Rev Bras Med do Esporte. 2005;11(1):81–5.

37. Marfell-Jones MJ, Stewart AD, de Ridder JH. International standards for anthropometric assessment. Wellington, New Zealand: International Society for the Advancement of Kinanthropometry; 2012.

38. Gore C. Physiological tests for elite athletes: Australian Sports Commission. Human Kinetics Publishers; 2000.

39. Bruton A, Conway J, Holgate S. Reliability: what is it, and how is it measured? Physiotherapy. 2000;86(2):94–9.

40. Weir J. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.

41. Stratford P, Agostino V, Brazeau C, Gowitzke B. Reliability of joint angle measurement: a discussion of methodology issues. Physiother Canada. 1984;36(1):5–9.

42. Myers J, Oyama S, Wassinger C, Ricci R, Abt J, Conley K, et al. Reliability, precision, accuracy, and validity of posterior shoulder tightness assessment in overhead athletes. Am J Sports Med. 2007;35(11):1922–30.

43. Portney L, Watkins M. Foundations of clinical research: applications to practice. Vol. 2. Prentice Hall Upper Saddle River; 2000.

44. Trippolini M, Dijkstra P, Geertzen J, Reneman M. Measurement properties of the Spinal Function Sort in patients with sub-acute whiplash-associated disorders. J Occup Rehabil. 2015;25(3):527–36.

45. Denegar C, Ball D. Assessing reliability and precision of measurement: an introduction to intraclass correlation and standard error of measurement. J Sport Rehabil. 1993;2(1):35–42.

46. Goto R, Mascie-Taylor C. Precision of measurement as a component of human variation. J Physiol Anthropol. 2007;26(2):253–6.

47. Bland M. An introduction to medical statistics. Oxford University Press; 1987.

48. Chinn S. Repeatability and method comparison. Thorax. 1991;46(6):454–6.

49. Atkinson G, Nevill A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sport Med. 1998;26(4):217–38.

50. Mueller W, Martorell R. Reliability and accuracy of measurement. In: Lohman T, Roche A, Martorell R, editors. Anthropometric standardisation reference manual. Champaign, IL: Human Kinetics Books; 1988. p. 83–6.

51. Wang Z, Zhong YQ, Chen KJ, Ruan JY, Zhu JC. 3D human body data acquisition and fit evaluation of

clothing. In: Advanced Materials Research. Trans Tech Publ; 2014. p. 4161–4.

52. Xu B, Yu W, Yao M, Pepper MR, Freeland-Graves JH. Three-dimensional surface imaging system for assessing human obesity. Opt Eng. 2009;48(10):107204.

53. ISO 7250. Basic human body measurements for technological design - Part 1: Body measurement definition and landmarks. Geneve, Switzerland: International Organization for Standardization; 2008.

54. ISO 20685. 3-D scanning methodologies for internationally compatible anthropometric databases. 2010.